

## Ejercicios Tema 3

REDES NEURONALES Y APRENDIZAJE PROFUNDO - IRINA ARÉVALO

- [1] Dado el texto2 de NLTK, estandariza, elimina las stopwords y la puntuación, lematiza y tokeniza el texto.
- [2] Estandariza, tokeniza y elimina las stopwords y la puntuación del siguiente texto:

texto = “Muchos años después, frente al pelotón de fusilamiento, el coronel Aureliano Buendía había de recordar aquella tarde remota en que su padre lo llevó a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y cañabrava construidas a la orilla de un río de aguas diáfanas que se precipitaban por un lecho de piedras pulidas, blancas y enormes como huevos prehistóricos. El mundo era tan reciente, que muchas cosas carecían de nombre, y para mencionarlas había que señalarlas con el dedo.”

- [3] Busca cómo encontrar los nombres de usuario de los siguientes correos usando regex:  
texto = “The new registrations are hansolo709@gmail.com , leia.mor101@gmail.com. If you find any disruptions, kindly contact luke111@gmail.com or yoda77@gmail.com ”
- [4] Extrae las 10 palabras más comunes que no sean stopwords ni puntuación del texto4 de NLTK.
- [5] Extrae los nombres propios de este texto:  
texto = “James works at Microsoft. She lives in manchester and likes to play the flute”
- [6] Repite el lab 4.3 pero usando one-hot encoding como transformación del texto. ¿Qué diferencias hay?
- [7] Usa el dataset Rotten Tomatoes (<https://www.kaggle.com/datasets/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset>) para predecir, usando el review\_content (contenido de la crítica), el resultado de la crítica (review\_type) y si el crítico es top (top\_critic).
- [8] Usa el dataset Genre Classification (<https://www.kaggle.com/datasets/hijest/genre-classification-dataset-imdb>) para clasificar el género de una película en función de su descripción
- [9] Usa alguno de los otros modelos de generación de lenguaje incluidos en la librería Transformers para generar un texto a raíz de ”Hello, I’m a language model,”. Compáralo con GPT-2.
- [10] Compara el uso del reconocimiento de objetos de Transformers con el de YOLO con la imagen dog.jpg.
- [11] Usa la tarea table-question-answering para responder preguntas sobre el dataset Titanic.